

# **CREDIT CARD FRAUD DETECTION PROJECT REPORT**

## **Introduction**

In this project, my aim was to tackle the problem of credit card fraud, which is a major concern affecting financial institutions and cardholders worldwide. Reports from the Federal Trade Commission have highlighted credit card fraud as the most common form of identity theft reported by consumers in the United States. My main goal was to build a machine learning model that could accurately identify fraudulent credit card transactions.

## **Dataset and Algorithms**

For this project, I worked with a dataset comprising nearly 28,500 credit card transactions made by European cardholders in September 2013. The dataset was imbalanced, meaning there were only a small number of fraudulent transactions compared to the regular ones. Despite this challenge, I proceeded with my analysis and used two powerful machine learning algorithms - the Logical Regression Algorithm and the Random Forest Algorithm - to detect fraud.

## **Exploratory Data Analysis**

During my exploratory data analysis, I examined the characteristics of the dataset. I focused on the time of transactions, the transaction amounts, and whether the transactions were fraudulent or not. Some interesting patterns emerged, such as two peak periods during the day when credit card transactions were more frequent. I also observed that most transactions involved smaller amounts, with a few outliers involving significantly larger sums.

## **Data Processing**

To ensure the reliability of my model, I processed the data, taking into account the imbalanced nature of the dataset. I divided the dataset into training, validation, and testing sets, and I used a technique called

Random Undersampling to balance the data by matching the number of fraudulent transactions with regular ones. I also standardized the data using StandardScaler to create a normal distribution of transaction amounts and time values.

## **Modeling and Evaluation of Performance**

With the data prepared, I built and deployed my Logical Regression and Random Forest models. I chose these algorithms because they are known for their interpretability, ability to handle large datasets, and their capability to avoid overfitting.

To evaluate the performance of my models, I used various metrics, including accuracy, precision, recall, and the F1 score. Accuracy measures the fraction of correct predictions made by my model, but it can be misleading for imbalanced datasets. Recall indicates the percentage of correctly identified fraudulent transactions, while precision represents the percentage of predicted fraudulent transactions that were actually fraudulent. The F1 score combines both recall and precision to provide a more effective evaluation metric for imbalanced classes.

## **Conclusion**

This project can be extended to include real-time fraud detection using web applications, mobile apps, or other platforms.

Thank you for taking the time to explore the Credit Card Fraud Detection project! If you have any questions or feedback, feel free to reach out to me at [osuolalefolarin@gmail.com](mailto:osuolalefolarin@gmail.com)

GitHub Repository: <https://github.com/Folarinosuolale/Data-Science-Machine-Learning>